Appendix B: Example Risk-tiering Materials for Generative AI

B.1: Example Adverse Impacts

Table B.1: Example adverse impacts, adapted from NIST 800-30r1 Table H-2 [**?**].

Level	**Description**

Harm to Operations	• Inability to perform current missions/business functions.
	  – In a sufficiently timely manner.
	  – With sufficient confidence and/or correctness.
	  – Within planned resource constraints.
	• Inability, or limited ability, to perform missions/business functions in the future.
	  – Inability to restore missions/business functions.
	  – In a sufficiently timely manner.
	  – With sufficient confidence and/or correctness.
	  – Within planned resource constraints.
	• Harms (e.g., financial costs, sanctions) due to noncompliance.
	  – With applicable laws or regulations.
	  – With contractual requirements or other requirements in other binding agreements (e.g., liability).
	• Direct financial costs.
	• Reputational harms.
	  – Damage to trust relationships.
	  – Damage to image or reputation (and hence future or potential trust relationships).

Harm to Assets	• Damage to or loss of physical facilities. • Damage to or loss of information systems or networks. • Damage to or loss of information technology or equipment. • Damage to or loss of component parts or supplies. • Damage to or of loss of information assets. • Loss of intellectual property.
Harm to Individuals	• Injury or loss of life. • Physical or psychological mistreatment. • Identity theft. • Loss of personally identifiable information. • Damage to image or reputation. • Infringement of intellectual property rights. • Financial harm or loss of income.

Harm to Other Organizations	• Harms (e.g., financial costs, sanctions) due to noncompliance.
	  – With applicable laws or regulations.
	  – With contractual requirements or other requirements in other binding agreements (e.g., liability).
	• Direct financial costs.
	• Reputational harms.
	  – Damage to trust relationships.
	  – Damage to image or reputation (and hence future or potential trust relationships).

Harm to the Nation	• Damage to or incapacitation of critical infrastructure.
	• Loss of government continuity of operations.
	• Reputational harms.
	  – Damage to trust relationships with other governments or with nongovernmental entities.
	  – Damage to national reputation (and hence future or potential trust relationships).
	• Damage to current or future ability to achieve national objectives.
	  – Harm to national security.
	• Large-scale economic or workforce displacement.

B.2: Example Impact Descriptions

Table B.2: Example Impact level descriptions, adapted from NIST SP800-30r1 Appendix H, Table H-3 [**?**].

Qualitative Values	**Semi-Quantitative Values**		**Description**
Very High	96-100	10	An incident could be expected to have multiple severe or catastrophic adverse effects on organizational operations, organizational assets, individuals, other organizations, or the Nation.
High	80-95	8	An incident could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, individuals, other organizations, or the Nation. A severe or catastrophic adverse effect means that, for example, the incident might: (i) cause a severe degradation in or loss of mission capability to an extent and duration that the organization is not able to perform one or more of its primary functions; (ii) result in major damage to organizational assets; (iii) result in major financial loss; or (iv) result in severe or catastrophic harm to individuals involving loss of life or serious life-threatening injuries.
Moderate	21-79	5	An incident could be expected to have a serious adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation. A serious adverse effect means that, for example, the incident might: (i) cause a significant degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is significantly reduced; (ii) result in significant damage to organizational assets; (iii) result in significant financial loss; or (iv) result in significant harm to individuals that does not involve loss of life or serious life-threatening injuries.
Low	5-20	2	An incident could be expected to have a limited adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation. A limited adverse effect means that, for example, the incident might: (i) cause a degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is noticeably reduced; (ii) result in minor damage to organizational assets; (iii) result in minor financial loss; or (iv) result in minor harm to individuals.
Very Low	0-4	0	An incident could be expected to have a negligible adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation.

B.3: Example Likelihood Descriptions

Table B.3: Example likelihood levels, adapted from NIST SP800-30r1 Appendix G, Table G-3 [**?**].

Qualitative Values	**Semi-Quantitative Values**		**Description**
Very High	96-100	10	An incident is almost certain to occur; or occurs more than 100 times a year.
High	80-95	8	An incident is highly likely to occur; or occurs between 10-100 times a year.
Moderate	21-79	5	An incident is somewhat likely to occur; or occurs between 1-10 times a year.
Low	5-20	2	An incident is unlikely to occur; or occurs less than once a year, but more than once every 10 years.
Very Low	0-4	0	An incident is highly unlikely to occur; or occurs less than once every 10 years.

B.4: Example Risk Tiers

Table B.4: Example risk assessment matrix with 5 impact levels, 5 likelihood levels, and 5 risk tiers, adapted from NIST SP800-30r1 Appendix I, Table I-2 [**?**].

Likelihood	**Level of Impact**				
	Very Low	**Low**	**Moderate**	**High**	**Very High**
Very High	Very Low (Tier 5)	Low (Tier 4)	Moderate (Tier 3)	High (Tier 2)	Very High (Tier 1)
High	Very Low (Tier 5)	Low (Tier 4)	Moderate (Tier 3)	High (Tier 2)	Very High (Tier 1)
Moderate	Very Low (Tier 5)	Low (Tier 4)	Moderate (Tier 3)	Moderate (Tier 3)	High (Tier 2)
Low	Very Low (Tier 5)	Low (Tier 4)	Low (Tier 4)	Low (Tier 4)	Moderate (Tier 3)
Very Low	Very Low (Tier 5)	Very Low (Tier 5)	Very Low (Tier 5)	Low (Tier 4)	Low (Tier 4)

B.5: Example Risk Descriptions

Table B.5: Example risk descriptions, adapted from NIST SP800-30r1 Appendix I, Table I-3 [**?**] .

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	Very high risk means that an incident could be expected to have multiple severe or catastrophic adverse effects on organizational operations, organizational assets, individuals, other organizations, or the Nation.
High	80-95	8	High risk means that an incident could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, individuals, other organizations, or the Nation.
Moderate	21-79	5	Moderate risk means that an incident could be expected to have a serious adverse effect on organizational operations, organizational assets, individuals, other organizations, or the Nation.
Low	5-20	2	Low risk means that an incident could be expected to have a limited adverse effect on organizational operations, organizational assets, individuals, other organizations, or the Nation.
Very Low	0-4	0	Very low risk means that an incident could be expected to have a negligible adverse effect on organizational operations, organizational assets, individuals, other organizations, or the Nation.

B.6: Practical Risk-tiering Questions

B.6.1: Confabulation: How likely are system outputs to contain errors? What are the impacts if errors occur?

B.6.2: Dangerous and Violent Recommendations: How likely is the system to give dangerous or violent recommendations? What are the impacts if it does?

B.6.3: Data Privacy: How likely is someone to enter sensitive data into the system? What are the impacts if this occurs? Are standard data privacy controls applied to the system to mitigate potential adverse impacts?

B.6.4: Human-AI Configuration: How likely is someone to use the system incorrectly or abuse it? How likely is use for decision-making? What are the impacts of incorrect use or abuse? What are the impacts of invalid or unreliable decision-making?

B.6.5: Information Integrity: How likely is the system to generate deepfakes or mis or disinformation? At what scale? Are content provenance mechanisms applied to system outputs? What are the impacts of generating deepfakes or mis or disinformation? Without controls for content provenance?

B.6.6: Information Security: How likely are system resources to be breached or exfiltrated? How likely is the system to be used in the generation of phishing or malware content? What are the impacts in these cases? Are standard information security controls applied to the system to mitigate potential adverse impacts?

B.6.7: Intellectual Property: How likely are system outputs to contain other entities' intellectual property? What are the impacts if this occurs?

B.6.8: Toxicity, Bias, and Homogenization: How likely are system outputs to be biased, toxic, homogenizing or otherwise obscene? How likely are system outputs to be used as subsequent training inputs? What are the impacts of these scenarios? Are standard nondiscrimination controls applied to mitigate potential adverse impacts? Is the application accessible to all user groups? What are the impacts if the system is not accessible to all user groups?

B.6.9: Value Chain and Component Integration: Are contracts relating to the system reviewed for legal risks? Are standard acquisition/procurement controls applied to mitigate potential adverse impacts? Do vendors provide incident response with guaranteed response times? What are the impacts if these conditions are not met?

B.7: AI Risk Management Framework Actions Aligned to Risk Tiering

GOVERN 1.3, GOVERN 1.5, GOVERN 2.3, GOVERN 3.2, GOVERN 4.1, GOVERN 5.2, GOVERN 6.1, MANAGE 1.2, MANAGE 1.3, MANAGE 2.1, MANAGE 2.2, MANAGE 2.3, MANAGE 2.4, MANAGE 3.1, MANAGE 3.2, MANAGE 4.1, MAP 1.1, MAP 1.5, MEASURE 2.6

Usage Note: Materials in Appendix B can be used to create or update risk tiers or other risk assessment tools for GAI systems or applications as follows:

- Table B.1 can enable mapping of GAI risks and impacts.
- Table B.2 can enable quantification of impacts for risk tiering or risk assessment.
- Table B.3 can enable quantification of likelihood for risk tiering or risk assessment.
- Table B.4 presents an example of combining assessed impact and likelihood into risk tiers.
- Table B.5 presents example risk tiers with associated qualitative, semi-quantitative, and quantitative values for risk tiering or risk assessment.
- Subsection B.6 presents example questions for qualitative risk assessment.
- Subsection B.7 highlights subcategories to indicate alignment with the AI RMF.